



별첨 사본은 아래 출원의 원본과 동일함을 증명함.

This is to certify that the following application annexed hereto
is a true copy from the records of the Korean Intellectual
Property Office.

출원번호 : 10-2003-0021482
Application Number

출원년월일 : 2003년 04월 04일
Date of Application APR 04, 2003

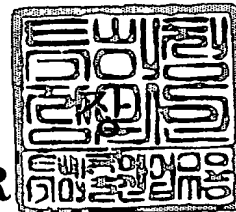
출원인 : 엔에이치엔(주)
Applicant(s) NHN Corporation

**PRIORITY
DOCUMENT**
SUBMITTED OR TRANSMITTED IN
COMPLIANCE WITH RULE 17.1(a) OR (b)



2004 년 03 월 25 일

특 허 청
COMMISSIONER



【서지사항】

【서류명】	특허출원서
【권리구분】	특허
【수신처】	특허청장
【참조번호】	0001
【제출일자】	2003.04.04
【발명의 명칭】	검색 엔진에서 등록된 웹사이트를 관리하기 위한 방법 및 그 시스템
【발명의 영문명칭】	A METHOD OF MANAGING WEB SITES REGISTERED IN SEARCH ENGINE AND A SYSTEM THEREOF
【출원인】	
【명칭】	엔에이치엔 (주)
【출원인코드】	1-1999-037388-8
【대리인】	
【성명】	천성진
【대리인코드】	9-2002-000294-5
【포괄위임등록번호】	2002-077078-8
【발명자】	
【성명의 국문표기】	나선균
【성명의 영문표기】	NA, Sun Kyoon
【주민등록번호】	791116-1065411
【우편번호】	138-151
【주소】	서울특별시 송파구 오륜동 올림픽선수촌아파트 303동 602호
【국적】	KR
【발명자】	
【성명의 국문표기】	이현정
【성명의 영문표기】	LEE, Hyun Jung
【주민등록번호】	730119-2058327
【우편번호】	152-092
【주소】	서울특별시 구로구 개봉2동 현대아파트 116동 1101호
【국적】	KR
【발명자】	
【성명의 국문표기】	최기정
【성명의 영문표기】	CHOI, Ki Jeong

【주민등록번호】 711223-1481512
【우편번호】 140-032
【주소】 서울특별시 용산구 이촌2동 209-22 시범아파트 7동 301호
【국적】 KR
【심사청구】 청구
【취지】 특허법 제42조의 규정에 의한 출원, 특허법 제60조의 규정에 의한 출원심사를 청구합니다. 대리인
천성진 (인)
【수수료】
【기본출원료】 20 면 29,000 원
【가산출원료】 25 면 25,000 원
【우선권주장료】 0 건 0 원
【심사청구료】 21 항 781,000 원
【합계】 835,000 원

【요약서】

【요약】

본 발명은 인터넷 상에서 소정의 웹사이트에 대한 정보를 제공하는 검색 엔진에 관한 것이다. 더욱 상세하게는, 검색 엔진에 등록되는 소정의 웹사이트에 대한 정보를 분석하여, 웹사이트에 포함하고 있는 콘텐츠의 주제와 상이한 검색 결과가 제공되지 않도록 등록된 웹사이트를 관리하기 위한 방법 및 그 시스템에 관한 것이다.

본 발명의 바람직한 일실시예에 따른 검색 엔진에서 웹사이트의 등록을 관리하기 위한 방법은 상기 등록된 웹사이트에 대한 정보를 수신하고, 상기 웹사이트 정보를 데이터베이스 수단에 소정의 필드 별로 분류하여 기록하는 단계; 검색 로봇을 제어하여 상기 등록된 웹사이트의 웹페이지를 구성하는 소스 파일을 독출하는 단계; 상기 독출된 소스 파일을 분석하는 단계; 소정의 기준에 따라 상기 웹사이트가 기만 사이트인지 여부를 판단하는 단계; 상기 웹사이트가 기만 사이트인 것으로 판단되는 경우, 상기 등록된 웹사이트에 대하여 소정의 처리를 수행하도록 제어하는 단계를 포함하는 것을 특징으로 하고, 바람직하게는 상기 소스 파일은 하이퍼텍스트 마크업 언어(HTML) 문서일 수 있다.

【대표도】

도 2

【색인어】

검색 엔진, 웹사이트, 등록, 기만 사이트, 키워드

【명세서】

【발명의 명칭】

검색 엔진에서 등록된 웹사이트를 관리하기 위한 방법 및 그 시스템{A METHOD OF MANAGING WEB SITES REGISTERED IN SEARCH ENGINE AND A SYSTEM THEREOF}

【도면의 간단한 설명】

도 1a는 웹사이트 검색 엔진 서비스를 제공하기 위한 종래의 시스템을 도시하는 구성 블록도이다.

도 1b는 웹사이트 검색 엔진 서비스 방식 중 디렉토리 검색 방식의 일예를 도시하는 도면이다.

도 2는 본 발명의 바람직한 일실시예에 따른 검색 엔진에서 등록된 웹사이트를 관리하기 위한 시스템을 도시하는 구성 블록도이다.

도 3은 본 발명의 일실시예에 따른 검색 엔진에서 등록된 웹사이트를 관리하기 위한 방법을 도시하는 흐름도이다.

도 4a 내지 도 4k는 본 발명의 바람직한 일실시예에 따른 검색 엔진에서 등록된 웹사이트를 관리하기 위한 방법에 있어서, 검색 로봇이 웹사이트를 순회하여 독출한 얻은 기만 사이트의 정보 유형들을 도시하는 도면이다.

도 5은 본 발명의 바람직한 일실시예에 따른 검색 엔진에서 등록된 웹사이트를 관리하기 위한 방법에 있어서, 기만 사이트로 판별된 웹사이트의 등록자에게 소정의 제재 조치를 가하는 방법을 도시하는 흐름도이다.

도 6은 본 발명에 따른 검색 엔진에서 웹사이트의 등록을 관리하는데 이용될 수 있는 범용 컴퓨터 시스템의 내부 블록도이다.

<도면의 주요 부분에 대한 부호의 설명>

201 : 인터페이스 모듈 202 : 웹사이트 등록 모듈
 203 : 웹사이트 관리 모듈 204 : 웹사이트 정보 DB
 205 : 웹사이트 분석 모듈 207 : 검색 로봇
 208 : 메일 서버 209 : SMS 서버

【발명의 상세한 설명】

【발명의 목적】

【발명이 속하는 기술분야 및 그 분야의 종래기술】

<13> 본 발명은 인터넷 상에서 소정의 웹사이트에 대한 정보를 제공하는 검색 엔진에 관한 것이다. 더욱 상세하게는, 검색 엔진에 등록되어 있는 소정의 웹사이트에 대한 정보를 분석하여, 웹사이트에 포함되어 있는 실제 콘텐츠와 상이한 검색 결과가 제공되지 않도록 등록된 웹사이트를 관리하기 위한 방법에 관한 것이다.

<14> 알타비스타(<http://www.altavista.com>), 라이코스(<http://www.lycos.com>), 야후(<http://www.yahoo.com>) 등과 같은 통상의 검색 엔진은 통상 웹사이트 정보를 소정의 기준에 따라 분류 및 저장하고 관리하기 위한 데이터베이스, 웹 상을 지속적으로 순회하면서 새로운 웹사이트 정보를 기계적으로 수집하기 위한 소프트웨어로 구현되는 검색 로봇(robot), 수집된 데이터를 데이터베이스화하여 상기 검색 엔진을 이용하는 이용자로 하여금 검색할 수 있도록 하는 검색 엔진 소프트웨어로 구성된다.

<15> 상술한 검색 엔진 서비스를 제공하기 위한 전체 시스템의 블록도가 도 1a에 도시되어 있다. 도 1a을 참조하면, 사용자는 사용자 단말기(110)를 통하여 인터넷을 통해 검색 엔진 서버(150)로 접속한다. 사용자가 소정의 검색어를 입력하면, 검색 엔진 서버(150)는 검색 엔진 소프트웨어(140)로 해당 검색어에 대한 웹사이트 정보를 쿼리(query)하고, 검색 엔진 소프트웨어(140)는 해당 데이터베이스(130)를 검색하여 소정의 웹사이트 정보를 알려 준다. 검색 로봇(120)은 상기 설명한 바와 같이, 웹 상을 지속적으로 순회하면서 웹 서버(160)로부터 새로운 웹사이트 정보를 기계적으로 수집하기 위한 소프트웨어로 구현되는 엔티티(entity)이다. 검색 로봇(120)은 네트워크 상에서 HTML(HyperText Markup Language)로 기술된 문장을 탐색하고, 기재되어 있는 링크처를 파싱(parsing)하여 네트워크 상에 존재하는 다수의 웹사이트로부터 데이터를 수집한다. 이와 같이 검색 로봇(120)에 의해 수집된 데이터는 데이터베이스화되는데, 여기에서 데이터베이스화라 함은 웹사이트에 위치하는 소정의 정보에 대해 형태소 분석(morphological analysis)을 수행하고, 인덱스 테이블을 작성하여 데이터베이스(130)에 저장하는 일련의 수순을 의미한다. 데이터베이스(130)는 검색 로봇(120)에 의해 수집된 모든 웹사이트 정보를 저장하기 위한 것이다. 검색 엔진 소프트웨어(140)는 검색 결과를 사용자에게 보여주는 기능을 한다. 이 소프트웨어는 데이터베이스(130)에 저장된 수많은 페이지를 검색하여 검색 결과물을 검색 용어와 일치되는 정확도의 순서로 나열하게 된다. 위와 같은 종래의 검색 엔진은 다음과 같은 방법으로 웹사이트에 대한 정보를 검색 엔진에 등록하고, 상기 정보를 사용자에게 제공한다.

<16> (1) 상술한 바와 같이 검색 로봇을 이용하여 소정의 정보를 수집하고, 수집된 정보를 전문 서퍼(surfer)의 검수를 거쳐 상기 웹사이트를 검색 엔진에 등록한다.

- <17> (2) 등록하고자 하는 웹사이트의 주제에 따라 분류된 디렉토리를 선택하고, 상기 선택된 디렉토리에 대해 상기 웹사이트 등록 신청을 하며, 전문 서퍼의 검수를 거쳐 검색 엔진에 등록된다. 검색 엔진에 따라 이러한 디렉토리 등록의 경우에는 소정의 등록료를 받고 등록에 소요되는 시간을 줄여주는 서비스를 제공하기도 한다.
- <18> 상기의 방법 등으로 검색 엔진에 등록된 웹사이트는 소정의 정보를 검색하고자 하는 사용자의 검색어 입력에 따라, 통합 웹 검색 또는 디렉토리 검색 등의 다양한 검색 방식에 따라 검색되어 사용자에게 제공된다. 상기 통합 웹 검색은 다른 용어로 "단어별 검색"이라고도 하는데, 이 검색 방법은 데이터베이스에 모든 웹사이트들의 범용 리소스 로케이터(URL)를 저장하고 특정 키워드(Keyword)를 입력함으로써 원하는 정보를 찾는 방법을 말한다. 또한, 상기 디렉토리 검색은 다른 용어로 "주제별 검색"이라고도 하는데, 이 검색 방법은 각 분야별로 분류가 되어 있고, 원하는 분야를 링크하게 되면 그 분야의 세부 항목을 상세히 볼 수 있도록 하는 검색 방법이다. 이와 같이 사용자가 계속 링크해서 세부 항목을 검색하는 경우 사용자가 원하는 정보를 찾을 수 있게 된다. 예를 들어, 2002년 한국 월드컵에서 한국팀의 경기 스코어를 찾고자 하는 경우, 스포츠 -> 구기종목 -> 축구 -> 월드컵 -> 한국/일본2002년 월드컵 -> 한국팀 경기 스코어 등의 방법으로 검색을 할 수 있게 된다. 도 1b에는 이러한 디렉토리 검색 방식의 일예가 출력된 화면이 도시되어 있다. "월드컵"에 대해 검색된 디렉토리로 "월드컵", "2002 FIFA 월드컵 한국 일본", "월드컵의 역사" 등의 디렉토리가 출력되고, 사용자는 상기 디렉토리 중 자신이 검색하고자 하는 정보가 위치할 가능성이 높은 디렉토리로 이동함으로써 정보 검색을 수행할 수 있다. 상술한 웹 통합 검색 방법을 사용하는 대표적인 검색 엔진으로는 카네기멜론 대학의 마이크 L. 몰딘이 개발한 라이코스(<http://lycos.cs.cmu.edu>) 검색 엔진을 들 수 있고, 상기 디렉토리 검색 방법을 사용하는 대표적인 검색 엔진으로는 야후

(http://www.yahoo.com)를 들 수 있다. 현재 다수의 검색 엔진들은 상술한 검색 방법을 함께 서비스 하는 하이브리드(hybrid) 방식의 서비스를 제공하고 있다.

<19> 상술한 종래 기술에 의한 검색 엔진에의 웹사이트 등록 방법과 등록된 웹사이트의 검색 방법에서는 다음과 같은 문제점이 있다.

<20> 인터넷 사용자가 급증하면서 소정의 정보를 검색하고자 하는 사용자의 수가 늘어나고, 그들이 검색하고자 하는 정보의 종류가 많아 지게 된다. 이러한 사용자 및 검색하고자 하는 정보의 종류의 증가에 따라, 출현 빈도가 높은 검색어가 생기게 되고 이러한 출현 빈도가 높은 검색어(이하, "인기 키워드"라고 함)를 자신의 웹사이트에 다양한 방법으로 삽입함으로써 상기 검색어를 이용하여 검색을 하고자 하는 사용자에게 전혀 무용한 콘텐츠가 포함된 웹사이트(이하, "기만 사이트"라고 함)에 대한 정보를 제공하게 되는 문제점이 있다. 예를 들어, 인기 키워드 중의 하나인 "피카츄"에 대한 정보를 검색하고자 하는 사용자가 "피카츄"를 검색어로 입력한 경우, 웹사이트에 "피카츄"가 포함되어 있는 모든 등록된 웹사이트 정보가 상기 사용자에게 제공되는데, 웹사이트들 중에서 콘텐츠는 성인물에 관한 것이면서, 웹사이트 중간 중간에 다양한(대부분의 경우 악의적인) 방법으로 "피카츄" 텍스트를 삽입한 웹사이트가 존재할 수 있고, 이러한 성인물을 콘텐츠로 하고 있는 웹사이트의 정보가 상술한 인기 키워드의 삽입으로 인해 다양한 연령대의 사용자에게 노출될 위험이 있다는 문제점이 있다.

<21> 상술한 문제점들을 해결하기 위한 해결책으로서 사용자들의 고발 신고 또는 전문 서퍼 등의 전문 인력을 통한 등록 웹사이트의 지속적인 모니터링이 필요하지만, 이러한 종래 기술에 따른 해결 방법은 상술한 문제점들에 대한 궁극적인 해결책이 될 수 없음은 자명하고, 이러한 문제점을 인터넷 상에서 소정의 알고리즘을 통하여 자동적으로 수행될 수 있는 방법이 도출될 수 있다면 상술한 문제점을 일거에 해결할 수 있는 유용한 수단이 될 수 있을 것이다.

【발명이 이루고자 하는 기술적 과제】

- <22> 본 발명에 따른 검색 엔진에서 웹사이트의 등록을 관리하기 위한 방법은 상술한 종래 기술의 문제점을 해결하기 위한 것으로서, 상술한 기만 사이트를 소정의 알고리즘을 이용하여 자동적으로 검출해 낼 수 있도록 함으로써 검색 엔진 사용자로 하여금 자신이 검색하고자 하는 정보를 정확히 검색할 수 있는 검색 엔진을 제공하는 것을 그 목적으로 한다.
- <23> 또한, 본 발명에 따른 검색 엔진에서 웹사이트의 등록을 관리하기 위한 방법은 상기 기만 사이트를 자동적으로 검출하고, 검출된 기만 사이트 운영자에 대한 제재 조치를 가하도록 함으로써 검색 엔진에 등록되는 웹사이트 자체의 자정이 강화될 수 있도록 하는 것을 그 목적으로 한다.
- <24> 또한, 본 발명에 따른 검색 엔진에서 웹사이트의 등록을 관리하기 위한 방법은 상기 기만 사이트의 검출 및 검출된 상기 사이트들에 대한 경고 등의 제재 조치를 소정의 알고리즘에 의해 자동적으로 수행되도록 함으로써, 상술한 기만 사이트 검출을 위해 소요될 수 있는 다수의 인력 자원을 절약할 수 있도록 하는 것을 그 목적으로 한다.

【발명의 구성 및 작용】

- <25> 본 발명의 바람직한 일실시예에 따른 검색 엔진에서 웹사이트의 등록을 관리하기 위한 방법은 상기 등록된 웹사이트에 대한 정보를 수신하고, 상기 웹사이트 정보를 데이터베이스 수단에 소정의 필드 별로 분류하여 기록하는 단계; 검색 로봇을 제어하여 상기 등록된 웹사이트의 웹페이지를 구성하는 소스 파일을 독출하는 단계; 상기 독출된 소스 파일을 분석하는 단계; 소정의 기준에 따라 상기 웹사이트가 기만 사이트인지 여부를 판단하는 단계; 상기 웹사이트가 기만 사이트인 것으로 판단되는 경우, 상기 등록된 웹사이트에 대하여 소정의 처리를 수행하

도록 제어하는 단계를 포함하는 것을 특징으로 하고, 바람직하게는 상기 소스 파일은 하이퍼텍스트 마크업 언어(HTML) 문서일 수 있다.

<26> 또한, 본 발명의 바람직한 일실시예에 따른 검색 엔진에서 등록된 웹사이트를 관리하기 위한 시스템은 하나 이상의 단말기와 데이터 통신을 수행하기 위한 인터페이스 모듈; 상기 하나 이상의 단말기로부터 소정의 웹사이트의 정보를 포함하는 웹사이트 등록 신청을 수신하고, 상기 웹사이트 정보를 소정의 필드 별로 분류하기 위한 웹사이트 등록 모듈; 상기 웹사이트 정보 및 상기 웹사이트에 대응하는 소정의 키워드를 분류하여 저장하기 위한 데이터베이스 수단; 상기 웹사이트의 웹페이지를 구성하는 소스 파일을 추출하고, 추출된 소스 파일을 분석하기 위한 웹사이트 분석 모듈; 및 소정의 기준에 따라 상기 웹사이트가 기만 사이트인지 여부를 판단하기 위한 웹사이트 관리 모듈을 포함하는 것을 특징으로 한다.

<27> 상술한 바와 같이, 본 명세서에서 사용되는 기만 사이트라 함은 웹페이지의 소스 파일 등에 다양한 방법으로 소정의 키워드 등을 삽입하여 검색어를 통해 검색되는 내용과 실제 웹사이트에 포함된 콘텐츠가 완전히 상이한 웹사이트를 의미한다. 본 발명의 일실시예에 의하면 상기 웹페이지의 소스 파일 등에 삽입되는 소정의 키워드는 인기 키워드일 수 있다.

<28> 또한, 본 명세서에서 사용되는 인기 키워드라 함은 인터넷 사용자가 입력하는 검색어 중 그 출현 빈도가 매우 높은 검색어를 의미하는데, 이러한 인기 키워드는 그 당시의 사회 상황과 인터넷 사용자들의 취향에 따라 지속적으로 변화할 수 있다. 이러한 인기 키워드에는 사회적으로 유해한 내용을 내포하고 있는 일종의 유해 키워드가 포함될 수 있는데, 이러한 유해 키워드의 예로는 "자살", "왕따", "도박", "범죄 모의" 등을 들 수 있다.

<29> 이하 첨부된 도면을 참조하여 본 발명의 바람직한 일실시예에 따른 검색 엔진에서 웹사이트의 등록을 관리하는 방법에 대하여 상술한다.

<30> 도 2는 본 발명의 바람직한 일실시예에 따른 검색 엔진에서 웹사이트의 등록을 관리하기 위한 시스템을 도시한 구성 블록도이다. 도2를 참조하면, 본 발명의 바람직한 일실시예에 따른 검색 엔진에서 웹사이트의 등록을 관리하기 위한 시스템은 인터페이스 모듈(201), 웹사이트 등록 모듈(202), 웹사이트 관리 모듈(203), 웹사이트 정보 데이터베이스(204), 웹사이트 분석 모듈(205), 및 검색 로봇(207)으로 구성될 수 있다. 또한, 본 발명의 바람직한 일실시예에 의하면 검색 엔진에서 웹사이트의 등록을 관리하기 위한 시스템은 등록 웹사이트의 등록자에게 소정의 메시지를 발송하기 위한 메일 서버(208), 또는 SMS 서버(209)를 포함할 수 있다. 이러한 메일 서버(208)와 SMS 서버(209)는 검색 엔진 서비스 제공 시스템 내에 포함될 수도 있고, 제3자가 운영하는 시스템 내에 위치할 수도 있다. 또한, 도 2에는 인터페이스 모듈(201), 다종의 모듈들, 및 메일 서버(208) 또는 SMS 서버(209)가 별개의 엔티티인 것처럼 도시되어 있으나 이는 설명의 편의를 위한 것일 뿐, 동일한 엔티티일 수 있다. 또한, 상기 도 2에 도시된 구성 요소들은 물리적으로도 동일한 장소에 위치할 수도 있고, 다른 실시예에 따르면 물리적으로 이격되어 있을 수도 있다.

<31> 먼저, 인터페이스 모듈(201)은 소정의 웹사이트를 검색 엔진에 등록하고자 하는 등록자 측에 구비된 컴퓨터 단말기와 검색 엔진의 등록 관리 시스템 사이의 데이터 전송 및 물리적 전송 장비 간의 인터페이스 역할을 담당하는 모듈이다.

<32> 웹사이트 등록 모듈(202)은 상기 등록자로부터 소정의 웹사이트에 대한 등록 신청을 수신하고, 웹사이트 등록 신청에 포함된 웹사이트에 대한 정보 데이터를 수집 및 분류하는 기능을 담당한다. 이러한 웹사이트 등록 모듈(202)은 웹사이트 등록에 대한 소정의 과금을 수행하

는 과금 모듈(도시되지 아니함)을 더 포함할 수 있고, 과금 모듈은 등록을 원하는 웹사이트의 종류(일반적인 내용을 담고 있는 일반 사이트 또는 성인 콘텐츠를 담고 있는 성인 사이트)에 따라 그 과금 내역을 달리 적용하도록 동작할 수 있다.

- <33> 웹사이트 관리 모듈(203)은 본 발명에 따른 웹사이트의 등록 관리를 총괄하는 모듈로서, 검색 로봇(207)이 수집한 웹사이트에 대한 정보를 기초로 상기 웹사이트가 선정된 기준에 적합하게 운영되고 있는지를 판단하고, 비정상적으로 운영되는 웹사이트, 즉 기만 사이트인 것으로 판단되는 경우, 상기 등록자에 대해 소정의 조치를 자동적으로 취하도록 제어하는 기능을 담당한다. 또한, 웹사이트 관리 모듈(203)은 메일 서버(208)나 단문자 메시지 서비스(SMS) 서버(209)와 연동함으로써 상기 기만 사이트의 등록자에 대해 이메일을 발송하거나 상기 등록자의 이동통신단말기로 SMS를 전송함으로써 웹사이트의 부정 운영에 대한 경고를 할 수 있다.
- <34> 웹사이트 정보 데이터베이스(204)는 등록된 웹사이트에 대한 정보를 분류하여 기록하는 역할을 담당한다. 웹사이트 정보 데이터베이스(204)에는 웹사이트의 범용 리소스 로케이터(URL), 해당 웹사이트의 키워드, 해당 웹사이트의 등록자 정보(등록자 이름, 주소, 이메일 주소, 이동통신단말기 번호 등), 해당 웹사이트의 디렉토리 정보 등의 다양한 정보가 각 필드 별로 분류 및 저장되어 있을 수 있다.
- <35> 본 발명에 따른 웹사이트 정보 데이터베이스(204)에 저장된 정보는 시스템 관리자 및 상기 웹사이트의 등록자에 의해서 수정될 수 있고, 웹사이트의 콘텐츠가 바뀌는 경우 등록자가 직접 수정하지 아니하더라도 검색 로봇(207)이 수집한 데이터의 분석 결과(해당 웹사이트의 URL에 대응하는 새로운 키워드 등) 등에 따라 자동적으로 해당 정보가 갱신되도록 동작할 수 있다.

- <36> 웹사이트 분석 모듈(205)은 검색 로봇(207)이 수집한 웹사이트의 정보를 분석하는 역할을 담당한다. 검색 로봇(207)이 수집하는 데이터의 종류와 분석 방법에 대해서는 아래 도 3에 대한 설명과 함께 상술한다.
- <37> 상술한 본 발명의 일실시예에 따른 검색 엔진에서 웹사이트의 등록을 관리하기 위한 시스템을 구성하는 각 구성 요소들은 설명의 편의를 위하여 단순히 기능적으로 구분된 것일 뿐, 각 구성 요소의 실제 물리적 위치와는 무관하다. 또한, 상술한 모듈들은 하드웨어로 구현될 수도 있고, 특정 코드를 이용한 소프트웨어로서 구현될 수 있음은 당업자에게 자명하다.
- <38> 도3은 본 발명의 바람직한 일실시예에 따른 검색 엔진에서 웹사이트의 등록을 관리하기 위한 방법을 도시하는 흐름도이다. 이하 도 4a 내지 도 4k, 및 도 6을 참조하여 도 3에 도시된 본 발명의 바람직한 일실시예에 따른 검색 엔진에서 웹사이트의 등록을 관리하기 위한 방법에 대해 상술한다.
- <39> 도 3에 도시된 본 발명의 바람직한 일실시예에 따른 검색 엔진에서 웹사이트의 등록을 관리하기 위한 방법은 다음과 같이 수행된다. 소정의 웹사이트를 검색 엔진에 등록하고자 하는 등록자는 등록을 원하는 웹사이트에 대한 정보와 함께 웹사이트 등록 신청을 한다(단계 305). 상기 웹사이트에 대한 정보는 웹사이트 정보 데이터베이스에 정보 필드((등록자 이름, 주소, 이메일 주소, 이동통신단말기 번호 등) 별로 분류되어 기록되고(단계 310), 상기 웹사이트는 검색 엔진에 등록된다(단계 315). 이러한 등록 단계(단계 315)는 몇 개의 루트를 통해 수행될 수 있는데, 먼저 상술한 것과 같이 웹사이트 관리자가 검색 엔진에 자신의 웹사이트의 등록을 의뢰함으로써 등록되는 경우가 있고, 검색 로봇이 웹 상을 무작위로 돌아다니면서 얻어온 웹사이트 정보를 이용하여 웹사이트가 검색 엔진에 등록되는 경우가 있을 수 있다. 전자의 경우에는 웹사이트 등록자 자신이 웹사이트의 주제(예를 들면, "피카츄", "변리사 시험" 등)

를 정하여 상기 웹사이트의 주제와 가장 근접한 카테고리에 웹사이트의 등록을 신청하고, 신청된 웹사이트에 대해서는 전문 서퍼의 검수를 거쳐 소정의 조건(웹사이트의 완성도, 등록 비용을 지불하지 아니하는 경우에는 비상업적 사이트 요건 충족 여부 등)을 만족하는 것으로 판단되는 경우에 검색 엔진에 등록될 수 있다. 본 발명에 따른 검색 엔진에서 웹사이트의 등록을 관리하기 위한 방법에서는 검색 엔진에 등록되는 루트를 상기 등록자의 신청에 의한 경우로 한정하여 설명하고 있지만, 본 발명에 따른 검색 엔진에서 웹사이트의 등록을 관리하기 위한 방법 및 그 시스템은 검색 엔진에 웹사이트가 등록되는 다종 다양한 방법에 대해서도 동일하게 적용될 수 있다.

- <40> 웹사이트가 등록되면, 검색 엔진은 검색 로봇을 제어하여 등록된 웹사이트의 웹페이지를 구성하는 소스 파일을 독출하고, 독출된 소스 파일을 분석한다(단계 320).
- <41> 본 발명의 일실시예에 따른 분석 방법은 하이퍼텍스트 마크업 언어(HyperText Markup Language; HTML) 문서를 분석하는 방법이다. 더욱 상세하게는 웹사이트의 HTML 문서 내의 태그를 분석함으로써, 출현 빈도가 높은 검색어, 즉 인기 키워드를 자신의 웹사이트를 구성하는 HTML 문서에 삽입한 웹사이트, 즉 기만 사이트인지 여부를 판단할 수 있게 된다. 당업자라면 주지하는 바와 같이, HTML 문서는 태그라는 일종의 명령어와 함께 구성되고, 웹사이트를 만드는 웹디자이너 등은 이러한 태그를 통하여 웹사이트를 구성하고, 자신의 웹사이트에서 제공하고자 하는 콘텐츠를 자신의 웹사이트에 포함시킨다.
- <42> 도 4a 내지 도 4k는 도 3a의 단계 320에서 수행되는, 웹사이트의 하이퍼텍스트 마크업 언어 문서를 분석하여 HTML 문서에 포함된 태그에 기만 사이트로 판단할 수 있는 소지가 있는 문자열을 포함하는 다양한 실시예들을 도시하는 도면으로서, 더욱 상세하게는 웹사이트의 HTML 문서 태그를 분석하여 다양한 방식의 기만 사이트를 검출하는 다양한 유형을 도시한 도면이다.

이하, 도 4a 내지 도 4k를 참조하여 본 발명에 따른 검색 엔진에서 웹사이트의 등록을 관리하기 위한 방법에서 HTML 문서의 분석이 어떻게 수행되는지에 대해 상술한다.

<43> (1) 배경색과 동일한 색의 문자열을 사용한 기만 사이트

<44> 도 4a는 웹사이트의 배경색과 동일한 색의 문자열을 태그로서 포함하고 있는 기만 사이트의 일예를 도시한 도면이다. 좌측의 도면은 사용자에게 디스플레이되는 웹사이트의 화면을 도시한 것이고, 우측의 도면은 좌측에 디스플레이되는 웹사이트의 HTML 소스 파일이다. 도 4a의 상단에 도시된 소스 파일을 보면, 백그라운드 컬러(bgcolor)로 #FFFFFF가 지정되어 있고 텍스트의 컬러도 #FFFFFF로 지정되어 있어서, 텍스트인 "스타크래프트"와 "줄라맨"이 웹사이트의 화면에는 보이지 않게 된다. 도 4a의 하단에 도시된 일예도 마찬가지로 백그라운드 컬러로 백색을 의미하는 #FFFFFF가 지정되어 있고 텍스트의 컬러도 백색(white)으로 지정되어 있어서, 텍스트인 "스타크래프트"와 "줄라맨"이 웹사이트의 화면에는 보이지 않게 된다. 도 4a의 소스 파일에 나타난 <body> 태그는 당업자라면 주지하는 바와 같이 웹브라우저에 표시되는 배경 또는 텍스트의 여러 속성을 설정할 수 있도록 하는 것이다. 태그는 크게 시작 태그와 종료 태그로 구성되는 태그(도 4a에 도시된 <body></body> 또는 태그)와 종료 태그의 필요가 없는 단독 태그로 구분될 수 있고, 이러한 태그를 이용하여 웹사이트를 다양한 방식으로 구성할 수 있다. 상기와 같이 배경색과 문자열의 색이 동일하다면, 이러한 웹사이트는 소정의 인기 키워드에 영합하여 상기 인기 키워드와 무관한 콘텐츠를 포함하고 있음에도 불구하고, 검색 결과 화면에 디스플레이될 수 있다.

<45> (2) 리다이렉션(redirection) 페이지에 포함된 문자열을 이용한 기만 사이트

<46> 도 4b는 리다이렉션 페이지에 포함된 문자열을 이용하는 기만 사이트의 일예를 도시한 도면이다. 좌측의 도면은 사용자에게 디스플레이되는 웹사이트의 화면을 도시한 것이고, 우측의 도

면은 좌측에 디스플레이되는 웹사이트의 HTML 소스 파일이다. 당업자라면 주지하는 바와 같이 리다이렉션은 접속된 웹사이트에서 새로운 웹사이트로의 이동을 설정하기 위한 것으로서, 도 4b의 우측에 도시된 소스 파일의 형태로 구현될 수 있다. 우측의 소스 파일 중 상단의 메타 태그를 이용한 일예를 보면, 메타 태그 중 http-equiv 속성을 이용하고 있다. 상기 메타 태그는 통상 정해진 시간(도 4b의 content 항목에서 지정되는 시간) 경과 후에 자동으로 다른 문서로 이동하도록 설정하기 위한 것으로서, 주로 홈페이지의 주소가 변경되는 경우 옛 주소로 접속한 사용자에게 주소 변경 안내를 보여 주고 소정의 시간 경과시 자동으로 새로운 주소로 이동할 수 있도록 하는데 사용된다. 또한, 도 4b 우측 중단 및 하단의 경우에도 각각 self.location 태그 및 location.replace 태그 등을 이용하여 현재의 웹페이지를 "http://www.naver.com"으로 리다이렉션하도록 동작한다.

<47> 도 4b에 도시된 리다이렉션 페이지를 이용한 기만 사이트의 일예에서, 우측 상단의 메타 태그의 경우 리다이렉션 명령 다음에, 중단 및 하단의 경우에는 </script> 태그 다음에 소정의 인기 키워드("스타크래프트", "줄라맨")를 삽입하고 있다.

<48> 이러한 리다이렉션 페이지의 경우에는 태그 자체가 다른 웹사이트의 이동을 명령하고 있는 것이므로, 태그 이후에 부가되는 텍스트는 아무런 역할을 하지 아니하는 대신, 검색 로봇의 검색 결과는 웹사이트 내에서 문자열 출현 빈도에 따라 정해지므로 상기 웹사이트의 주제를 원 주제와 다르게 판단할 수 있다. 따라서, 상기와 같이 리다이렉션 페이지에 문자열이 포함되어 있다면, 이러한 웹사이트는 인기 키워드에 영합하여 상기 인기 키워드와 무관한 콘텐츠를 포함하고 있음에도 불구하고, 검색 결과 화면에 디스플레이될 수 있다.

<49> (3) 타이틀 태그를 이용한 문자열을 이용한 기만 사이트

<50> 도 4c는 타이틀 태그에 포함되는 문자열을 이용하는 기만 사이트의 일예를 도시한 도면이다.

좌측의 도면은 사용자에게 디스플레이되는 웹사이트의 화면을 도시한 것이고, 우측의 도면은 좌측에 디스플레이되는 웹사이트의 HTML 소스 파일이다. 당업자라면 주지하는 바와 같이 타이틀 태그는 웹브라우저의 상단에 웹사이트의 주제를 간략하게 디스플레이하기 위하여 사용하는 태그로서, 도 4c의 우측에 도시된 소스 파일의 형태로 구현될 수 있다. 우측의 소스 파일 중 상단의 타이틀 태그를 이용한 일예를 보면, 타이틀 태그 내에 "스타크래프트"와 "줄라맨" 등의 인기 키워드를 다수 포함하고 있고, 이 태그로 인해 좌측의 웹브라우저와 같이 출력된다. 또한, 도 4c의 하단 도면에 도시된 경우에는 복수 개의 타이틀 태그를 사용하고 있다. 하단 도면의 소스 파일에는 <title> 및 </title> 태그 사이에 "히딩크, "스타크래프트", "줄라맨" 등의 다수의 인기 키워드가 포함되어 있다.

<51> 이러한 타이틀 태그의 경우에는 타이틀 태그 내에 아무리 많은 문자열을 포함하고 있더라도 웹브라우저에는 디스플레이되지 않는 반면, 검색 로봇의 검색 결과는 웹사이트 내에서 문자열 출현 빈도에 따라 정해지므로 상기 타이틀 태그에 포함된 문자열들로 말미암아 상기 웹사이트의 주제를 원 주제와 다르게 판단할 수 있다. 따라서, 상기와 같이 타이틀 태그에 포함된 문자열의 길이가 소정치 이상이거나 타이틀 태그의 개수가 복수 개라면, 이러한 웹사이트는 인기 키워드에 영합하여 상기 인기 키워드와 무관한 콘텐츠를 포함하고 있음에도 불구하고, 검색 결과 화면에 디스플레이될 수 있다.

<52> (4) 메타 태그(meta tag)에 포함되는 문자열을 이용한 기만 사이트

<53> 도 4d는 메타 태그에 포함된 문자열을 이용하는 기만 사이트의 일예를 도시한 도면이다.

좌측의 도면은 사용자에게 디스플레이되는 웹사이트의 화면을 도시한 것이고, 우측의 도면은 좌측에 디스플레이되는 웹사이트의 HTML 소스 파일이다.

<54> 당업자라면 주지하는 바와 같이 메타 태그는 해당 HTML 문서의 작성자, 작성 날짜, 키워드 등 웹브라우저의 본문에 디스플레이되지 아니하는 HTML 문서에 대한 일반적인 정보를 표현하고자 할 경우 사용하는 것이다. 도 4d의 우측 소스 파일을 보면, 메타 태그 내에 문서 이름(name)은 description이고, 문서의 내용(content)은 "스타크래프트"와 "줄라맨" 등의 인기 키워드를 다수 포함하고 있다. 이 경우, 메타 태그에 포함된 상기 인기 키워드 중인 문자열은 웹브라우저에 표시되지 아니하는 반면, 검색 로봇의 검색 결과는 웹사이트 내에서 문자열 출현 빈도에 따라 정해지므로 상기 웹사이트의 주제를 원 주제와 다르게 판단할 수 있다. 따라서, 상기와 같이 메타 태그 내에 문자열을 포함하고 있고 상기 문자열의 길이가 소정치 이상이라면, 이러한 웹사이트는 인기 키워드에 영합하여 상기 인기 키워드와 무관한 콘텐츠를 포함하고 있음에도 불구하고, 검색 결과 화면에 디스플레이될 수 있다.

<55> (5) 프레임 태그에 위치한 문자열을 이용한 기만 사이트

<56> 도 4e는 프레임 태그에 위치한 문자열을 이용하는 기만 사이트의 일예를 도시한 도면이다. 좌측의 도면은 사용자에게 디스플레이되는 웹사이트의 화면을 도시한 것이고, 우측의 도면은 좌측에 디스플레이되는 웹사이트의 HTML 소스 파일이다. 당업자라면 주지하는 바와 같이 프레임 태그는 웹브라우저에 표시되는 화면을 둘 또는 그 이상으로 분할하고자 하는 경우 사용하는 것이다. 도 4e의 우측 소스 파일을 보면, 프레임 태그 <FRAMESET ROWS=" " >는 화면을 가로로 분할하기 위한 태그이고 " " 내에는 화면을 나누는 할당 정보가 삽입된다. 이러한 프레임 태그의 종료(</FRAMESET>) 태그 이후에 위치하는 문자열은 "스타크래프트"와 "줄라맨" 등의 인기 키워드를 다수 포함하고 있다. 이 경우, 프레임 태그의 종료 태그 이후에 위치하는 상기 인기 키워드 중인 문자열은 웹브라우저의 화면 분할과는 아무런 관계가 없는 반면, 검색 로봇의 검색 결과는 웹사이트 내에서 문자열 출현 빈도에 따라 정해지므로 상기 웹사이트의 주

제를 원래의 주제와 다르게 판단할 수 있다. 따라서, 상기와 같이 프레임 태그에 문자열이 위치하고 상기 문자열의 길이가 소정치 이상이라면, 이러한 웹사이트는 인기 키워드에 영합하여 상기 인기 키워드와 무관한 콘텐츠를 포함하고 있음에도 불구하고, 검색 결과 화면에 디스플레이될 수 있다.

<57> (6) 폼(form) 태그에 포함된 문자열을 이용한 기만 사이트

<58> 도 4f는 폼 태그에 포함된 문자열을 이용하는 기만 사이트의 일예를 도시한 도면이다.

좌측의 도면은 사용자에게 디스플레이되는 웹사이트의 화면을 도시한 것이고, 우측의 도면은 좌측에 디스플레이되는 웹사이트의 HTML 소스 파일이다. 당업자라면 주지하는 바와 같이 폼 태그는 웹브라우저에 출력되는 소정의 폼을 정의하기 위해 사용하는 것이다. 도 4f의 우측 소스 파일을 보면, 폼 태그는 "<form> <input type="버튼종류" value="보여지는 글자"></form>"의 형식으로 구성될 수 있다. 따라서, 도 4f의 우측에 도시된 소스 파일의 경우는 버튼 종류가 "hidden"이므로 버튼에 어떠한 글자도 보여질 수 없도록 설정된 것이고, 보여지는 글자 즉, 실제로는 웹브라우저에 표시되지 아니하는 문자열에는 "스타크래프트"와 "줄라맨" 등의 인기 키워드가 다수 포함되어 있다. 이 경우, 폼 태그에 포함된 상기 인기 키워드 중인 문자열은 웹브라우저의 폼을 정의하는 것과는 아무런 관계가 없는 반면, 검색 로봇의 검색 결과는 웹사이트 내에서 문자열 출현 빈도에 따라 정해지므로 상기 웹사이트의 주제를 원 주제와 다르게 판단할 수 있다. 따라서, 상기와 같이 폼 태그에 포함된 문자열의 길이가 소정치 이상이라면, 이러한 웹사이트는 인기 키워드에 영합하여 상기 인기 키워드와 무관한 콘텐츠를 포함하고 있음에도 불구하고, 검색 결과 화면에 디스플레이될 수 있다.

<59> (7) Div 태그에 포함된 문자열을 이용한 기만 사이트

<60> 도 4g는 div 태그에 포함된 문자열을 이용하는 기만 사이트의 일예를 도시한 도면이다. 좌측의 도면은 사용자에게 디스플레이되는 웹사이트의 화면을 도시한 것이고, 우측의 도면은 좌측에 디스플레이되는 웹사이트의 HTML 소스 파일이다. 당업자라면 주지하는 바와 같이 div 태그는 통상 ID와 클래스 속성을 사용하여 스타일 시트와 함께 사용하는 것이다. 도 4g의 우측 소스 파일을 보면, div 태그는 "<div style='display:none; ...'"으로 기재되어 있다. 이 경우 웹브라우저에 출력되는 문자열의 속성(style)이 디스플레이(display) 없음(none)이므로 이후에 등장하는 문자열은 웹브라우저에 표시되지 아니한다, 이 경우, div 태그에 포함된 상기 인기 키워드 중인 문자열은 웹브라우저의 화면 디스플레이와는 아무런 관계가 없는 반면, 검색 로봇의 검색 결과는 웹사이트 내에서 문자열 출현 빈도에 따라 정해지므로 상기 웹사이트의 주제를 원 주제와 다르게 판단할 수 있다. 따라서, 상기와 같이 div 태그에 포함된 문자열의 길이가 소정치 이상인 경우, 이러한 웹사이트는 인기 키워드에 영합하여 상기 인기 키워드와 무관한 콘텐츠를 포함하고 있음에도 불구하고, 검색 결과 화면에 디스플레이될 수 있다.

<61> (8) a href 태그에 포함된 문자열을 이용한 기만 사이트

<62> 도 4h는 a href 태그에 포함된 문자열을 이용하는 기만 사이트의 일예를 도시한 도면이다. 좌측의 도면은 사용자에게 디스플레이되는 웹사이트의 화면을 도시한 것이고, 우측의 도면은 좌측에 디스플레이되는 웹사이트의 HTML 소스 파일이다. 당업자라면 주지하는 바와 같이 a href 태그는 한 문서 내에서 또는 다른 문서나 타 웹사이트로 쉽게 이동하기 위해서 문서 내에 특정 글자나 이미지에 이동할 주소를 연결하기 위하여 사용하는 것이다. 도 4h의 우측 소스 파일을 보면, a href 태그는 "<a href='#이동할 위치 또는 주소'" 링크표시대상 "의 형식으로 구성될 수 있다. 도 4h의 우측에 도시된 소스 파일을 참조하면, 이동할 위치 및 링크 표시 대상이 전혀 지정되어 있지 아니하므로, 이러한 a href 태그는 웹브라우저 상에서 표

시는 물론 실행도 되지 아니하는 태그이다. 이러한 실행되지 아니하는 태그에 포함된 문자열에 "스타크래프트"와 "줄라맨" 등의 인기 키워드가 다수 포함되어 있다. 이 경우, a href 태그에 포함된 상기 인기 키워드 중인 문자열은 웹브라우저의 화면 출력 또는 링크 실행과는 아무런 관계가 없는 반면, 검색 로봇의 검색 결과는 웹사이트 내에서 문자열 출현 빈도에 따라 정해지므로 상기 웹사이트의 주제를 원 주제와 다르게 판단할 수 있다. 따라서, 상기와 같이 a href 태그에 포함된 문자열의 길이가 소정치 이상인 경우, 이러한 웹사이트는 인기 키워드에 영합하여 상기 인기 키워드와 무관한 콘텐츠를 포함하고 있음에도 불구하고, 검색 결과 화면에 디스플레이될 수 있는 위험이 있다.

<63> (9) 링크 팜(link farm)을 이용한 기만 사이트

<64> 도 4i는 a href 태그에 포함된 문자열을 이용하는 기만 사이트의 일예를 도시한 도면이다. 당업자라면 주지하는 바와 같이, 링크 팜은 자신의 웹사이트 내의 웹페이지들을 상호 링크하여 검색 로봇으로 하여금 계속하여 상기 웹사이트를 검색하도록 함으로써 궁극적으로 웹페이지의 순위를 높이기 위한 수단으로 사용되는 경우가 많다. 이러한 링크 팜은 상술한 a href 태그를 이용하여 구현될 수 있다.

<65> 이러한 링크 팜을 이용하는 웹사이트는 직접적으로 기만 사이트로 판단하기는 곤란하지만, 링크 팜을 일정 개수 이상으로 과다하게 사용하는 경우라면 검색 엔진으로 하여금 웹페이지 내의 인기 키워드들을 지속적으로 검색하도록 하는 방법을 사용하는 경우에는 기만 사이트일 가능성이 높으므로 이를 검출하여야 할 필요가 있다.

<66> (10) font 태그에 포함된 문자열을 이용한 기만 사이트

- <67> 도 4j는 font 태그에 포함된 문자열을 이용하는 기만 사이트의 일예를 도시한 도면이다. 좌측의 도면은 사용자에게 디스플레이되는 웹사이트의 화면을 도시한 것이고, 우측의 도면은 좌측에 디스플레이되는 웹사이트의 HTML 소스 파일이다.
- <68> 당업자라면 주지하는 바와 같이 font 태그는 웹브라우저에 표시되는 문자열의 크기 등을 지정하기 위해 사용되는 것으로서, 도 4j에 도시된 소스 파일에는 폰트 크기(font size)가 0으로 설정되어 있다. 따라서, 이러한 font 태그에 포함된 문자열은 웹브라우저 상에서 전혀 표시가 되지 아니한다. 이와 같이 폰트 크기가 0이어서 표시되지 아니하는 문자열에 "스타크 래프트"와 "줄라맨" 등의 인기 키워드가 다수 포함되어 있는 경우, font 태그에 포함된 상기 인기 키워드 등인 문자열은 웹브라우저의 화면 출력과는 아무런 관계가 없는 반면, 검색 로봇의 검색 결과는 웹사이트 내에서 문자열 출현 빈도에 따라 정해지므로 상기 웹사이트의 주제를 원 주제와 다르게 판단할 수 있다. 따라서, 상기와 같이 font 태그에 포함된 문자열의 크기가 0인 경우, 이러한 웹사이트는 인기 키워드에 영합하여 상기 인기 키워드와 무관한 콘텐츠를 포함하고 있음에도 불구하고, 검색 결과 화면에 디스플레이될 수 있다.
- <69> (11) 이미지 태그에 포함된 문자열을 이용한 기만 사이트
- <70> 도 4k는 img 태그에 포함된 문자열을 이용하는 기만 사이트의 일예를 도시한 도면이다. 좌측의 도면은 사용자에게 디스플레이되는 웹사이트의 화면을 도시한 것이고, 우측의 도면은 좌측에 디스플레이되는 웹사이트의 HTML 소스 파일이다.
- <71> 당업자라면 주지하는 바와 같이 img 태그는 문서에 소정의 이미지를 삽입하기 위하여 사용되는 것으로서, 도 4k에 도시된 소스 파일에는 삽입하고자 하는 이미지 파일이 "a.gif"로 지정되어 있다. img 태그에서는 삽입될 이미지가 지정된 다음, 상기 이미지의 위치, 정렬 방식 등의 속성을 지정하게 되는데, 도 4k의 경우에는 이러한 속성을 문자열로서 지정하고 있다.

이 경우, 웹브라우저 상에 상기 이미지가 디스플레이되는 경우 문자열로 지정된 속성은 상기 이미지 디스플레이에 어떠한 영향도 주지 못한다. 이와 같이 이미지의 속성에 영향을 미치지 아니하는 문자열에 "스타크래프트"와 "줄라맨" 등의 인기 키워드가 다수 포함되어 있는 경우, img 태그에 포함된 상기 인기 키워드 중인 문자열은 웹브라우저의 화면 출력과는 아무런 관계가 없는 반면, 검색 로봇의 검색 결과는 웹사이트 내에서 문자열 출현 빈도에 따라 정해지므로 상기 웹사이트의 주제를 원 주제와 다르게 판단할 수 있다. 따라서, 상기와 같이 img 태그에 포함된 문자열의 길이가 소정치 이상인 경우, 이러한 웹사이트는 인기 키워드에 영합하여 상기 인기 키워드와 무관한 콘텐츠를 포함하고 있음에도 불구하고, 검색 결과 화면에 디스플레이될 수 있다.

<72> 단계 320는 상술한 실시예와 같이 HTML 문서 내에 포함된 태그 등을 분석하여 태그 등에 기재된 문자열의 길이 등을 측정하는 단계이다. 이러한 측정 결과를 토대로 단계 325에서 측정 결과를 토대로 소정의 기준에 따라 상기 웹사이트가 기만 사이트인지 여부를 판단하게 된다.

<73> 단계 325에서 분석된 웹사이트가 기만 사이트인지 여부를 판단하기 위한 소정의 기준의 예는 도 4a 내지 도 4k에서 상술한 바와 같다. 일례로 상기 소정의 기준은 상기 하이퍼텍스트 마크업 언어 문서 중에 상기 웹페이지의 배경색과 동일한 색의 문자열이 포함되어 있는지 여부, 또는 상기 소정의 기준은 상기 하이퍼텍스트 마크업 언어 문서 중의 리다이렉션 (redirection) 태그에 문자열이 포함되어 있는지 여부일 수 있다.

<74> 본 발명의 바람직한 일실시예에 따르면, 단계 325의 소정의 기준은 상술한 (1) 내지 (11)의 기만 사이트 유형에서 설명한 분석 내용을 하이브리드(hybrid) 방식으로 적용하여 소정의 기준을 넘는 경우 기만 사이트로 판단하는 방법을 이용할 수 있다. 일례로, 타이틀 태그

내에 포함된 타이틀 문자열의 개수가 2개 이상인 경우부터 하나당 10점을 가산하고 최대 70점 까지 가산할 수 있다고 설정한다. 또한 리다이렉션 페이지에 포함된 문자열이 있는 경우 개수에 상관없이 70점을 가산하고, 링크 팜의 경우 링크 50개당 4 점씩 최대 80점 까지 가산하는 것으로 설정할 수 있다. 또한 크기가 0인 문자열이 있는 경우 문자열의 길이 100 byte 당 5점씩 가산하여 최대 70점 까지 가산하는 것으로 설정할 수 있다. 이와 같이 웹페이지를 구성하는 소스 파일을 분석하고, 상술한 여러 가지 기준에 따른 각각의 포인트와 포인트 가중치를 고려하여 계산한 총합계가 100점이 넘는 경우 해당 웹사이트는 기만 사이트라고 판단할 수 있을 것이다. 이러한 종합적인 판단 방법을 이용하는 경우, 하나의 유형에 근거한 판단(예를 들어, 타이틀 태그에 포함된 문자열의 개수가 50개이므로 기만 사이트로 판단)의 경우 그 판단에 오류가 있을 가능성이 높으므로, 상기와 같이 종합적으로 판단하여 기만 사이트 여부를 판단하는 것이 바람직할 수 있다.

<75> 또한, 본 발명의 바람직한 일실시예에 따르면, 상기 소정의 기준은 검색 로봇에 의한 웹 검색 방식에 의해 등록된 웹사이트와 소정의 디렉토리 지정 방식에 의해 등록된 웹사이트의 경우 차등적으로 적용할 수 있는데, 일예로 전자의 방식으로 등록된 웹사이트의 경우 웹페이지의 소스 파일 분석 결과 상술한 11가지 유형 중 3개의 유형에 해당하는 경우 기만 사이트로 판단한다면, 후자의 방식으로 등록된 웹사이트의 경우에는 1개의 유형에 해당하는 경우라도 기만 사이트로 판단하는 등의 방법을 사용할 수 있다. 이는 후자의 등록 방식의 경우 대부분의 검색 엔진 운영업체가 소정의 등록 비용을 웹사이트의 등록자로부터 받기 때문에 무료로 등록되는 전자의 경우보다는 어느 정도 호의를 베풀어야 할 필요가 있기 때문이다.

<76> 단계 325에서 기만 사이트로 판단된 경우, 상기 데이터베이스 수단의 웹사이트 등록자 필드를 검색하여 상기 웹사이트의 등록자 정보를 획득한다(단계 330). 상기 웹사이트의 등록

자 정보에서 상기 등록자의 연락 정보를 추출하고(단계 335), 추출된 연락 정보를 이용하여 상기 웹사이트의 등록자에 대해 이메일 발송 또는 단문자 메시지 전송 등의 소정의 경고 조치가 수행된다(단계 340). 이러한 경고 조치에 대해서는 후술하는 도 5에 대한 설명에서 상술한다.

<77> 본 발명의 또 다른 일실시예에 의하면, 단계 320에서 분석되는 것은 태그에 기재된 이미지일 수 있다. 예를 들어, 이미지의 화소를 분석하여 화소를 구성하는 RGB 콤포넌트를 추출하고, 추출 결과 특정 컬러(황색 등)의 화소가 소정의 기준(예를 들어 총 화소수의 50% 이상)을 초과하는 경우에는 일응 해당 사이트가 음란물 콘텐츠를 게시하고 있는 사이트일 것으로 추측할 수 있고, 이를 토대로 기만 사이트인지 여부를 판단할 수 있을 것이다.

<78> 도 5은 본 발명의 바람직한 일실시예에 따른 검색 엔진에서 웹사이트의 등록을 관리하기 위한 방법에 있어서, 기만 또는 변질 사이트로 판별된 웹사이트의 등록자에게 소정의 제재 조치를 가하는 방법을 도시하는 흐름도이다.

<79> 도 5를 참조하면, 상술한 도 3의 단계 325에서 기만 사이트로 판단된 경우의 자동 제재 조치가 도시되어 있다. 기만 사이트로 판단된 경우, 웹사이트 관리 모듈은 웹사이트 정보 데이터베이스를 검색하여 해당 웹사이트의 등록자의 정보를 획득하고(단계 510), 웹사이트 관리 모듈은 상기 등록자의 정보를 수신한다(단계 520 및 550). 본 발명의 일실시예에 따르면, 웹사이트 관리 모듈은 상기 등록자 정보에서 등록자의 이메일 주소 또는 이동통신단말기 번호 등의 연락 정보를 추출하고(단계 530), 메일 서버 또는 SMS 서버를 제어하여 상기 연락 정보로 소정의 메시지를 전송하도록 동작한다(단계 540).

<80> 본 발명의 또 다른 일실시예에 따르면, 웹사이트 관리 모듈은 상기 등록자 정보에서 등록자의 기타 등록 웹사이트 정보를 추출(단계 560)하고, 동일한 등록자 명의로 등록된 기타 웹사이트에 대한 분석(단계 570)을 자동적으로 수행하도록 제어한다. 동일 등록자 명의의 웹사

이트라면 동일 또는 유사한 방법으로 기만 사이트를 운영할 가능성이 높기 때문이다. 본 실시예의 경우, 기타 웹사이트의 분석 결과가 기만 사이트로 판단된 경우에는 도 5의 단계 510이 반복될 수 있다.

<81> 본 발명의 바람직한 일실시예에 의하면, 소정의 웹사이트가 상술한 분석 및 판단 방법에 의하여 기만 사이트로 판단된 경우, 상기 웹사이트의 등록자에게 자동적으로 이메일, 단문자 메시지 등을 발송하여 해당 웹사이트의 문제점을 지적하고 일정 유예 기간을 두어 시정을 요구하도록 동작할 수 있다. 또한, 일정 유예 기간 경과 후 자동적으로 상기 분석 및 판단 프로세스를 수행하도록 설정할 수 있고, 이 경우에도 시정이 안된 경우에는 등록 취소 등의 제재 조치를 취하는 것도 가능하다. 또한, 본 발명의 다른 일실시예에 의하면, 상기 웹사이트의 등록자에게는 추후 다른 웹페이지를 등록하고자 하는 경우 등록 절차를 까다롭게 하는 등의 제재를 가하는 것도 가능하다.

<82> 본 발명의 실시예들은 다양한 컴퓨터로 구현되는 동작을 수행하기 위한 프로그램 명령을 포함하는 컴퓨터 판독 가능 매체를 포함한다. 상기 컴퓨터 판독 가능 매체는 프로그램 명령, 데이터 파일, 데이터 구조 등을 단독으로 또는 조합하여 포함할 수 있다. 상기 매체는 프로그램 명령은 본 발명을 위하여 특별히 설계되고 구성된 것들이거나 컴퓨터 소프트웨어 당업자에게 공지되어 사용 가능한 것일 수도 있다. 컴퓨터 판독 가능 기록 매체의 예에는 하드 디스크, 플로피 디스크 및 자기 테이프와 같은 자기 매체(magnetic media), CD-ROM, DVD와 같은 광기록 매체(optical media), 플롭티컬 디스크(floptical disk)와 같은 자기-광 매체(magneto-optical media), 및 롬(ROM), 램(RAM), 플래시 메모리 등과 같은 프로그램 명령을 저장하고 수행하도록 특별히 구성된 하드웨어 장치가 포함된다. 상기 매체는 프로그램 명령, 데이터 구조 등을 지정하는 신호를 전송하는 반송파를 포함하는 광 또는 금속선, 도파관 등의 전

송 매체일 수도 있다. 프로그램 명령의 예에는 컴파일러에 의해 만들어지는 것과 같은 기계어 코드뿐만 아니라 인터프리터 등을 사용해서 컴퓨터에 의해서 실행될 수 있는 고급 언어 코드를 포함한다.

<83> 도 6은 본 발명에 따른 검색 엔진에서 등록된 웹페이지를 관리하는데 이용될 수 있는 범용 컴퓨터 시스템의 내부 블록도이다.

<84> 컴퓨터 시스템은 램(RAM: Random Access Memory)(660)과 롬(ROM: Read Only Memory)(670)을 포함하는 주기억장치와 연결되는 하나 이상의 프로세서(640)를 포함한다. 프로세서(640)는 중앙처리장치(CPU)로 불리기도 한다. 당업계에서 널리 알려져 있는 바와 같이, 롬(670)은 데이터(data)와 명령(instruction)을 단방향성으로 CPU에 전달하는 역할을 하며, 램(660)은 통상적으로 데이터와 명령을 양방향성으로 전달하는 데 사용된다. 램(660) 및 롬(670)은 컴퓨터 판독 가능 매체의 어떠한 적절한 형태를 포함할 수 있다. 대용량 기억장치(Mass Storage)(610)는 양방향성으로 프로세서(640)와 연결되어 추가적인 데이터 저장 능력을 제공하며, 상기된 컴퓨터 판독 가능 기록 매체 중 어떠한 것일 수 있다. 대용량 기억장치(610)는 프로그램, 데이터 등을 저장하는데 사용되며, 통상적으로 주기억장치보다 속도가 느린 하드디스크와 같은 보조기억장치이다. CD 롬(620)과 같은 특정 대용량 기억장치가 사용될 수도 있다. 프로세서(640)는 비디오 모니터, 트랙볼, 마우스, 키보드, 마이크로폰, 터치스크린형 디스플레이, 카드 판독기, 자기 또는 종이 테이프 판독기, 음성 또는 필기 인식기, 조이스틱, 또는 기타 공지된 컴퓨터 입출력장치와 같은 하나 이상의 입출력 인터페이스(630)와 연결된다. 마지막으로, 프로세서(640)는 네트워크 인터페이스(650)를 통하여 유선 또는 무선 통신 네트워크에 연결될 수 있다. 이러한 네트워크 연결을 통하여 상기된 방법의 절차를 수행할

수 있다. 상기된 장치 및 도구는 컴퓨터 하드웨어 및 소프트웨어 기술 분야의 당업자에게 잘 알려져 있다.

<85> 상기된 하드웨어 장치는 본 발명의 동작을 수행하기 위해 하나 이상의 소프트웨어 모듈로서 작동하도록 구성될 수도 있다.

【발명의 효과】

<86> 본 발명에 따른 검색 엔진에서 웹사이트의 등록을 관리하기 위한 방법에 의하면, 상술한 기만 사이트를 소정의 알고리즘을 이용하여 자동적으로 검출해 낼 수 있도록 함으로써 검색 엔진 사용자 하에금 자신이 검색하고자 하는 정보를 정확히 검색할 수 있는 검색 엔진을 제공할 수 있다는 기술적 효과를 얻을 수 있다.

<87> 또한, 본 발명에 따른 검색 엔진에서 웹사이트의 등록을 관리하기 위한 방법에 의하면, 상기 기만 사이트를 자동적으로 검출하고, 검출된 기만 사이트 운영자에 대한 제재 조치를 가하도록 함으로써 검색 엔진에 등록되는 웹사이트 자체의 자정 노력을 강화할 수 있다는 기술적 효과를 얻을 수 있다.

<88> 또한, 본 발명에 따른 검색 엔진에서 웹사이트의 등록을 관리하기 위한 방법에 의하면, 상기 기만 사이트의 검출 및 검출된 상기 사이트들에 대한 경고 등의 제재 조치를 소정의 알고리즘에 의해 자동적으로 수행되도록 함으로써, 상술한 기만 사이트 검출을 위해 소요되던 다수의 인력 자원을 절약할 수 있다는 기술적 효과를 얻을 수 있다.

<89> 이상과 같이 본 발명은 비록 한정된 실시예와 도면에 의해 설명되었으나, 본 발명은 상기의 실시예에 한정되는 것은 아니며, 이는 본 발명이 속하는 분야에서 통상의 지식을 가진 자라면 상기 기재로부터 다양한 수정 및 변형이 가능하다는 점은 자명하다. 따라서, 본 발명 사

상은 아래에 기재된 특허 청구 범위에 의해서만 파악되어야 하고, 이의 균등 또는 등가적 변형 모두는 본 발명 사상의 범주에 속한다고 할 것이다.

【특허청구범위】**【청구항 1】**

검색 엔진에서 등록된 웹사이트를 관리하기 위한 방법에 있어서,

상기 등록된 웹사이트에 대한 정보를 수신하고, 상기 웹사이트 정보를 데이터베이스 수단에 소정의 필드 별로 분류하여 기록하는 단계;

상기 등록된 웹사이트의 웹페이지를 구성하는 소스 파일을 독출하는 단계;

상기 독출된 소스 파일을 분석하는 단계;

소정의 기준에 따라 상기 웹사이트가 기만 사이트인지 여부를 판단하는 단계;

상기 웹사이트가 기만 사이트인 것으로 판단되는 경우, 상기 등록된 웹사이트에 대하여 소정의 처리를 수행하도록 제어하는 단계

를 포함하는 것을 특징으로 하는 등록된 웹사이트의 관리 방법.

【청구항 2】

제1항에 있어서,

상기 소스 파일은 하이퍼텍스트 마크업 언어(Hypertext Markup Language) 문서인 것을 특징으로 하는 등록된 웹사이트의 관리 방법.

【청구항 3】

제2항에 있어서,

상기 소정의 기준은 상기 하이퍼텍스트 마크업 언어 문서 중에 상기 웹페이지의 배경색과 동일한 색의 문자열이 포함되어 있는지 여부인 것을 특징으로 하는 등록된 웹사이트의 관리 방법.

【청구항 4】

제2항에 있어서,

상기 소정의 기준은 상기 하이퍼텍스트 마크업 언어 문서 중의 리다이렉션(redirection) 태그에 문자열이 포함되어 있는지 여부인 것을 특징으로 하는 등록된 웹사이트의 관리 방법.

【청구항 5】

제2항에 있어서,

상기 소정의 기준은 상기 하이퍼텍스트 마크업 언어 문서에 포함된 타이틀(title) 태그의 길이가 소정치 이상인지 여부 또는 타이틀 태그의 개수가 2개 이상인지 여부인 것을 특징으로 하는 등록된 웹사이트의 관리 방법.

【청구항 6】

제2항에 있어서,

상기 소정의 기준은 상기 하이퍼텍스트 마크업 언어 문서에 포함된 메타(meta) 태그 내의 문자열의 길이가 소정치 이상인지 여부인 것을 특징으로 하는 등록된 웹사이트의 관리 방법.

【청구항 7】

제2항에 있어서,

상기 소정의 기준은 상기 하이퍼텍스트 마크업 언어 문서 중의 프레임(frame) 태그에 문자열이 있는지 여부인 것을 특징으로 하는 등록된 웹사이트의 관리 방법.

【청구항 8】

제2항에 있어서,

상기 소정의 기준은 상기 하이퍼텍스트 마크업 언어 문서 중의 폼(form) 태그에 포함된 문자열의 길이가 소정치 이상인지 여부인 것을 특징으로 하는 등록된 웹사이트의 관리 방법.

【청구항 9】

제2항에 있어서,

상기 소정의 기준은 상기 하이퍼텍스트 마크업 언어 문서 중의 Div 태그에 동일한 문자열이 소정치 이상인지 여부인 것을 특징으로 하는 등록된 웹사이트의 관리 방법.

【청구항 10】

제2항에 있어서,

상기 소정의 기준은 상기 하이퍼텍스트 마크업 언어 문서 중의 A href 태그에 범용 리소스 로케이터가 아닌 문자열이 포함되어 있는지 여부인 것을 특징으로 하는 등록된 웹사이트의 관리 방법.

【청구항 11】

제2항에 있어서,

상기 소정의 기준은 상기 하이퍼텍스트 마크업 언어 문서에서 동일한 웹사이트 내의 웹 페이지를 링크하는 링크의 수가 소정의 개수 이상인지 여부인 것을 특징으로 하는 등록된 웹사이트의 관리 방법.

【청구항 12】

제2항에 있어서,

상기 소정의 기준은 상기 하이퍼텍스트 마크업 언어 문서에서 폰트(font)의 크기가 0인 문자열이 포함되어 있는지 여부인 것을 특징으로 하는 등록된 웹사이트의 관리 방법.

【청구항 13】

제2항에 있어서,

상기 소정의 기준은 상기 하이퍼텍스트 마크업 언어 문서 중의 Img 태그에 포함된 문자열의 길이가 소정치 이상인지 여부인 것을 특징으로 하는 등록된 웹사이트의 관리 방법.

【청구항 14】

제2항에 있어서,

상기 소정의 기준은 제3항 내지 제13항에 기재된 상기 기준들 중 적어도 두 개 이상을 포함하는 것을 특징으로 하는 등록된 웹사이트의 관리 방법.

【청구항 15】

제14항에 있어서,

상기 기만 사이트인지 여부를 판단하는 단계는,

상기 기준들에 따른 소정의 가중치를 유지하는 단계;

상기 기준들 각각에 따른 소정의 포인트 계산 방법에 따라 상기 기준들 각각에 대한 포인트를 계산하는 단계;

상기 기준들 각각에 대하여 계산된 상기 포인트와 상기 포인트에 대응하는 기준에 따른 가중치를 승산하여 상기 기준들 각각에 대한 중간값을 산정하는 단계;

상기 산정된 상기 기준들 각각에 대한 중간값을 합산하는 단계; 및

상기 합산된 중간값이 소정치 이상인지 여부를 판단하여 상기 소정치 이상인 경우 상기 웹사이트를 기만 사이트로 판단하는 단계

를 포함하는 것을 특징으로 하는 등록된 웹사이트의 관리 방법.

【청구항 16】

제2항에 있어서,

상기 소정의 기준은 상기 하이퍼텍스트 마크업 언어 문서 중의 Img 태그에 포함된 이미지 파일에 포함된 화소의 특정 RGB(Red, Green, Blue)의 조합이 소정의 값 이상인지 여부인 것을 특징으로 하는 등록된 웹사이트의 관리 방법.

【청구항 17】

제1항에 있어서,

상기 데이터베이스 수단은 웹사이트 등록자 필드를 포함하고,

상기 소정의 처리를 수행하도록 제어하는 단계는 ,

상기 데이터베이스 수단의 웹사이트 등록자 필드를 검색하여 상기 웹사이트의 등록자 정보를 획득하는 단계;

상기 웹사이트의 등록자 정보에서 상기 등록자의 연락 정보를 추출하는 단계; 및

추출된 상기 연락 정보로 메시지를 전송하는 단계

를 포함하는 것을 특징으로 하는 웹사이트의 등록 관리 방법.

【청구항 18】

제17항에 있어서,

상기 연락 정보는 상기 웹사이트 등록자의 이메일 주소 또는 이동통신단말기 번호인 것을 특징으로 하고, 상기 메시지를 전송하는 단계는

이메일 서버를 제어하여 상기 이메일 주소로 이메일을 발송하도록 하는 단계 또는 단문자 메시지 서비스 서버를 제어하여 상기 이동통신단말기 번호로 단문자 메시지를 발송하도록 하는 단계를 포함하는 것을 특징으로 하는 등록된 웹사이트의 관리 방법.

【청구항 19】

제1항에 있어서,

상기 데이터베이스 수단은 웹사이트 등록자 필드를 포함하고,

상기 소정의 처리를 수행하도록 제어하는 단계는,

상기 데이터베이스 수단의 웹사이트 등록자 필드를 검색하여 상기 웹사이트의 등록자 정보를 획득하는 단계;

상기 웹사이트의 등록자 정보에서 상기 등록자가 등록한 다른 웹사이트의 범용 리소스 로케이터(URL) 정보를 추출하는 단계; 및

상기 범용 리소스 로케이터(URL)를 통해 접속되는 웹사이트의 웹페이지를 구성하는 소스 파일을 독출하는 단계;

상기 독출된 소스 파일을 분석하는 단계;

소정의 기준에 따라 상기 웹사이트가 기만 사이트인지 여부를 판단하는 단계; 및

상기 웹사이트가 기만 사이트인 것으로 판단되는 경우, 상기 등록된 웹사이트에 대하여 소정의 처리를 수행하도록 제어하는 단계

를 포함하는 것을 특징으로 하는 등록된 웹사이트의 관리 방법.

【청구항 20】

제1항 내지 제19항 중 어느 한 항의 방법을 실행시키기 위한 프로그램을 기록한 컴퓨터 판독 가능한 기록 매체.

【청구항 21】

등록된 웹사이트를 관리하기 위한 시스템에 있어서,

하나 이상의 단말기와 데이터 통신을 수행하기 위한 인터페이스 모듈;

상기 하나 이상의 단말기로부터 소정의 웹사이트의 정보를 포함하는 웹사이트 등록 신청을 수신하고, 상기 웹사이트 정보를 소정의 필드 별로 분류하기 위한 웹사이트 등록 모듈;

상기 웹사이트 정보 및 상기 웹사이트에 대응하는 소정의 키워드를 분류하여 저장하기 위한 데이터베이스 수단;

상기 웹사이트의 웹페이지를 구성하는 소스 파일을 추출하고, 추출된 소스 파일을 분석하기 위한 웹사이트 분석 모듈; 및

소정의 기준에 따라 상기 웹사이트가 기만 사이트인지 여부를 판단하기 위한 웹사이트 관리 모듈

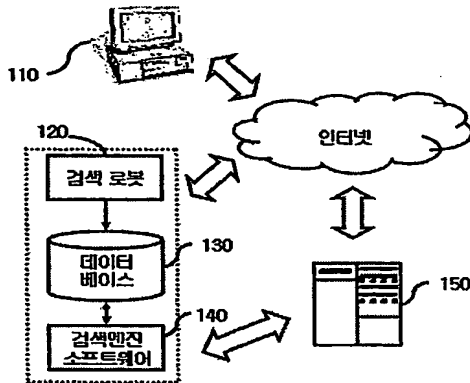
을 포함하는 것을 특징으로 하는 웹사이트 등록 관리 시스템.

102[REDACTED]21482

출력 일자: 2004/4/1

【도면】

【도 1a】



【도 1b】

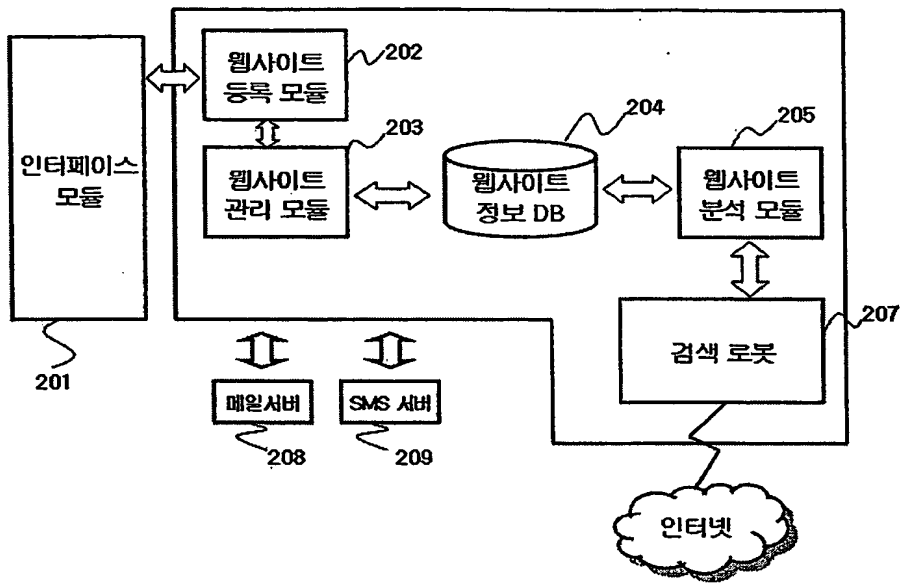
다독토리 검색 (2019 205대의 사이트를 찾아드립니다.)

<ul style="list-style-type: none"> 다 월드컵 다 신문동료 	<ul style="list-style-type: none"> 2002 FIFA 월드컵 한국 일본 스폰서, 협찬사 	<ul style="list-style-type: none"> 월드컵의 역사 축구선수 	<ul style="list-style-type: none"> 월드컵 경기장 전황이지 	<ul style="list-style-type: none"> 분석전술과 대표팀 뉴스와 미디어
---	---	---	---	--

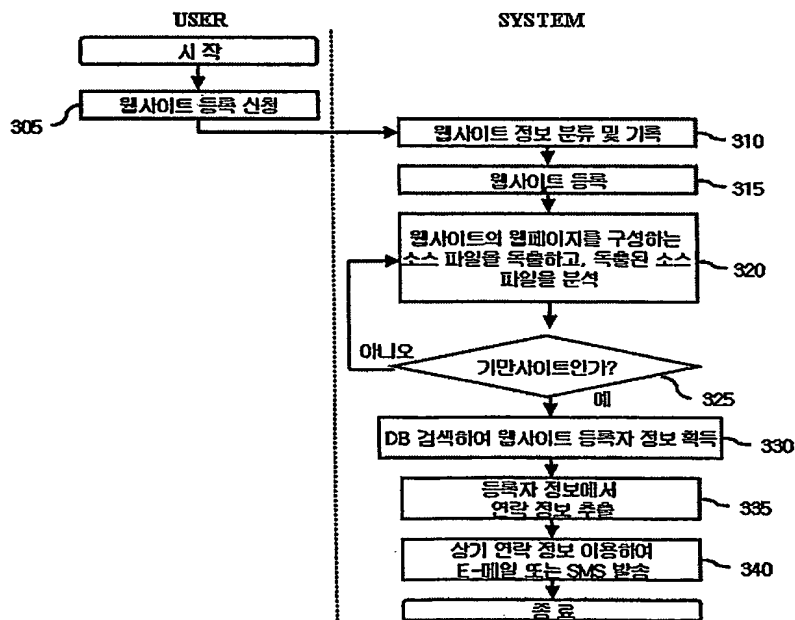
※ 웹사이트 : 관전순 보기 - 1 | 경기순 보기 - 1 | 선수순 보기

- 2002 FIFA 월드컵 공식홈 (한국판) - 에너스타사, 반올림, 한국의 알면의 스포 지평,
<http://fifa.worldcup.yahoo.com/ko/entry/>
 열린대륙으로 > 축구 > 대륙결산 > 대륙별 장문 > 축구
- 2002 FIFA 월드컵 한국/일본 TM (전황지) - 제화회사, 공식 마스코트, 경기 동영상, 역사, 국가대표팀, 분석, 선수관, 조추첨, 토너먼트 경기결과 및 경기, 축구계 소식, 경기장내 대화 구경 정보 지평, 공식 홈페이지,
<http://www.fifa-worldcup.com/>
 국가별관문 스포츠 > 스포츠 > 축구 > 경기 대회 > 월드컵 > 2002 FIFA 월드컵 한국 일본
- 2002 월드컵 한국조직위원회 (전황지) - 2002 FIFA 월드컵 대회 유치과정, 개최도시별 안내, 입장 전등록 조추첨, 지역별 행사 및 축제 대회 정보, 뉴스 자료, 경기 입장권 티켓 정보관 판매, 공식 홈페이지,
<http://www.2002worldcup.or.kr/>
 국가별관문 스포츠 > 스포츠 > 축구 > 경기 대회 > 월드컵 > 2002 FIFA 월드컵 한국 일본

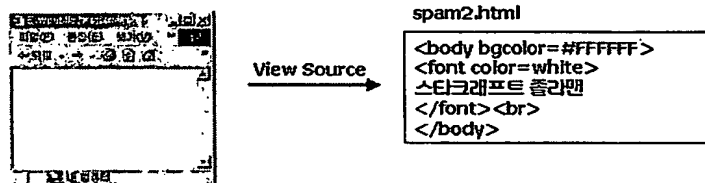
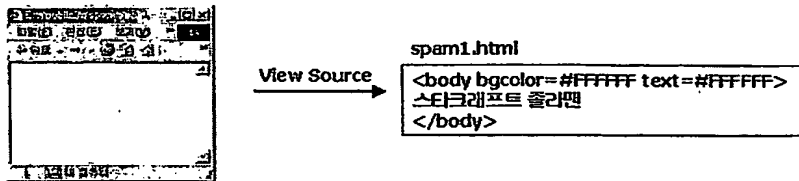
【도 2】



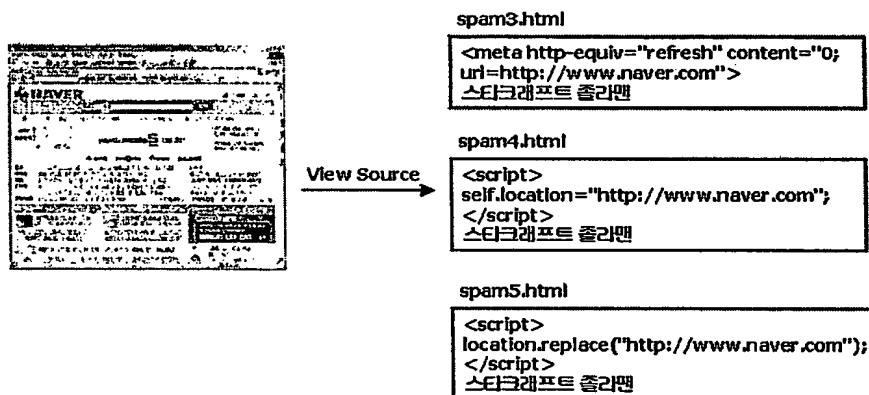
【도 3】



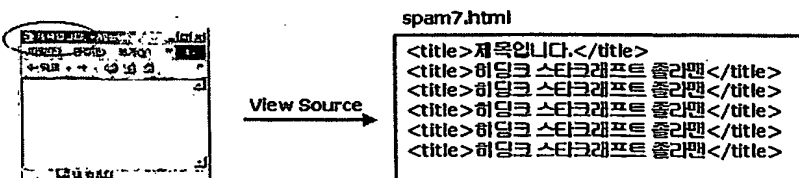
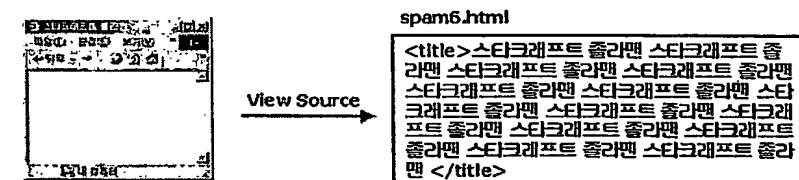
【도 4a】



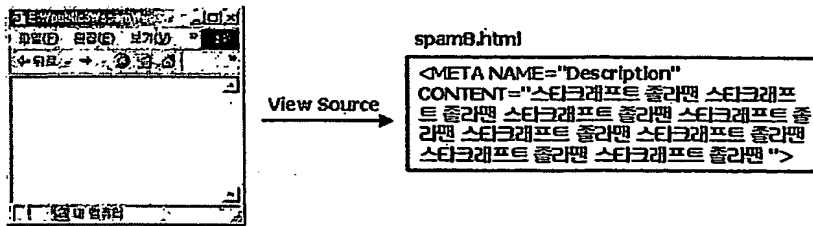
【도 4b】



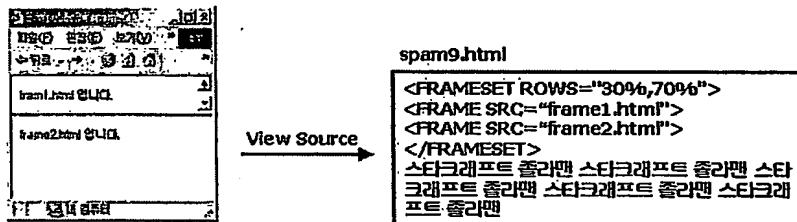
【도 4c】



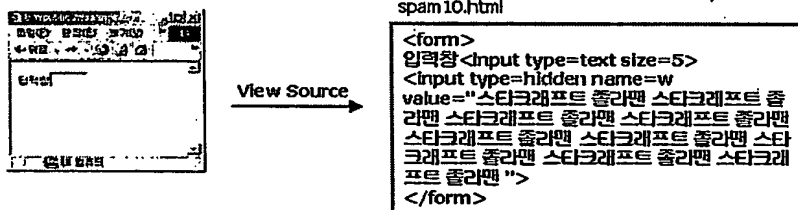
【도 4d】



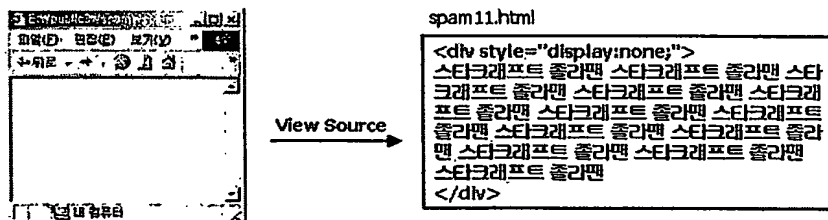
【도 4e】



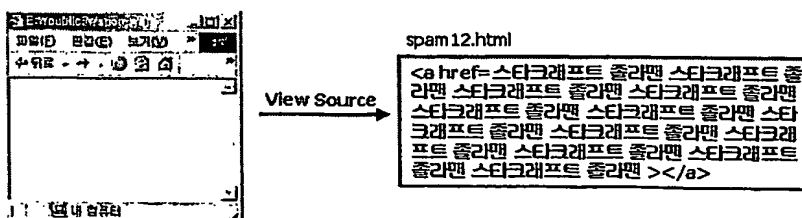
【도 4f】



【도 4g】



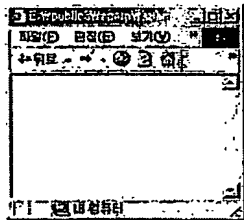
【도 4h】



【도 4i】

```
<a href="a11.htm">섹스</a><a href="a12.htm">야동</a><a href="a13.htm">야동
</a><a href="a14.htm">섹스</a><a href="a14.htm">야동</a><a href="a14.htm">
섹스</a><a href="a14.htm">섹스</a><a href="a14.htm">야동</a><a
href="a14.htm">야동</a>
```

【도 4j】

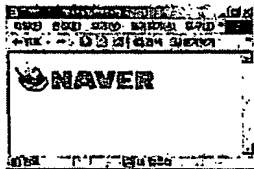


View Source

spam13.html

```
<font size=0>
스타크래프트 즐라맨 스타크래
프트 즐라맨 스타크래프트 즐
라맨 스타크래프트 즐라맨 스타
크래프트 즐라맨 스타크래프
트 즐라맨
</font>
```

【도 4k】

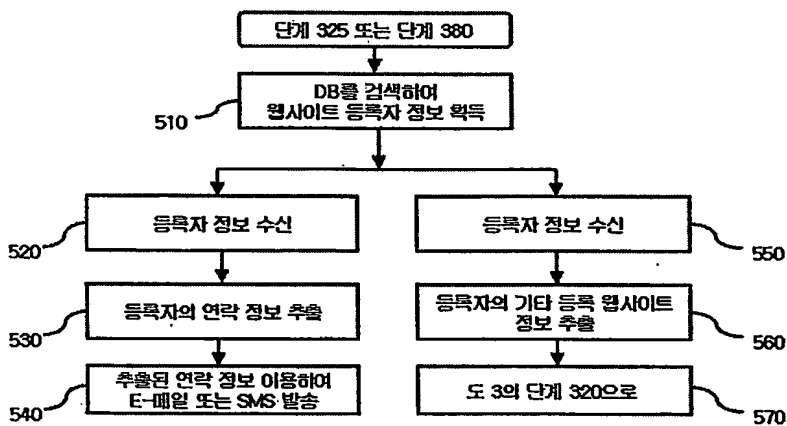


View Source

spam14.html

```
<img src=a.gif
alt="스타크래프트 즐라맨 스타크래프트
즐라맨 스타크래프트 즐라맨 스타크래프
트 즐라맨 스타크래프트 즐라맨 스타크래
프트 즐라맨 스타크래프트 즐라맨 스타크
래프트 즐라맨 스타크래프트 즐라맨 스타
크래프트 즐라맨 스타크래프트 즐라맨">
```

【도 5】



【도 6】

